Journal club & Exercise

Jinho
Postdoctoral Fellow
Linding lab, DTU


Protein kinase

Protein kinase is an enzyme that puts a phosphate on its target proteins (substrates).

**Phosphorylatable amino acids**

- Serine (Ser)
- Threonine (Thr)
- Tyrosine (Tyr)
Kinases in cellular signaling process

Writer-Reader-Eraser toolkit

Switch

www.examiner.com
Motif (word in natural language)

Kinases recognize specific sequence motifs and phosphorylate them. (Every kinase has its favorite words.)

eg)

![ATM Kinases](image)

Different kinases recognize different motifs. Some of them allow several amino acids at one position, which can be represented by position specific score matrix (PSSM).

eg)
NetPhorest pipeline

Linear Motif Atlas for Phosphorylation-Dependent Signaling

Martin Lee Miller,1,2* Lars Juhl Jensen,2,3* Francesca Diella,3 Claus Jørgensen,4
Michele Tinti,5 Lei Li,6 Marilyn Hsiung,4 Sirlester A. Parker,7 Jennifer Bordeaux,7
Thomas Sicheritz-Ponten,1 Marina Olhovsky,4 Adrian Pasculescu,4 Jes Alexander,8
Stefan Knapp,9 Nikolaj Blom,1 Peer Bork,2,10 Shawn Li,6 Gianni Cesareni,5 Tony Pawson,4
Benjamin E. Turk,7 Michael B. Yaffe,8† Søren Brunak,1,2† Rune Linding4,8,11†

(Published 2 September 2008)
A

Organization

- In vivo
- In vitro

In vivo phosphorylation sites → Phylogenetic trees → In vitro assays

B

Compilation

- Positive
- Negative
- Excluded

Extraction of positive and negative examples for each domain or family of related domains
Overview of the performance of the NetPhoreRest classifiers

Area under receiver operating characteristic curve
Systematic Discovery of In Vivo Phosphorylation Networks


2007
Sequence motifs lack sufficient information to uniquely identify the physiological substrates (Linding et al. 2007)

Cellular context:
- Subcellular compartmentalization
- Colocalization via anchoring protein & scaffolds
- Substrate capture by noncatalytic interaction domain
- Temporal & cell type specific coexpression
- Kinase-docking motifs within substrates
- Regulatory subunits
Cellular context - Subcellular compartmentalization

A kinase and a substrate should exist in the same place at the same time.

eg) Subcellular compartmentalization

http://bio3400.nicerweb.com
NetworKIN Algorithm

MS identification of phosphorylation sites

Rad50

Matching of sequence motifs for kinase families

Rad50

Construction of a context network from STRING

Rad50

- Scansite
- NetPhosK

Effects of Including Substrate Context
New NetworKIN Framework

**NetworKIN**

- MS identification of phosphorylation sites
- Manual annotation of phosphorylation sites
- Matching of sequence motifs for kinase families
- Construction of a context network from STRING

**NetPhorest**

- Protein domain sequences
- In vitro specificity assays
- Pipeline
- Motif atlas
- Protein domain sequences
- In vitro specificity assays
- Design of consensus antibodies
- Detection of purification biases

*Linding et al. Cell, 129,*

Miller et al. Science Signaling
2008 Sep 2;1(35):ra2.

http://NetworKIN.info

http://NetPhorest.info
KinomeXplorer is an integrated framework for modeling kinase-substrate interactions and aid in the design of inhibitor-based follow-up perturbation experiments. An interactive web interface allows investigation of predicted kinase-substrate interactions from human and major eukaryotic model organisms.

NetworKIN

NetworKIN is a method for predicting in vivo kinase-substrate relationships, that augments intrinsic specificities of kinases with cellular context for kinases and phosphoproteins.

NetPhorest

NetPhorest is a non-redundant collection of 222 sequence-based classifiers for linear motifs in phosphorylation-dependent signaling, including kinases, phosphatases, SH3 domains and more.

Additional resource: KinomeSelector which enables the user select an optimal kinase panel to assess the kinome-wide selectivity of kinase inhibitors.
NetworKIN is an integrated framework for modeling kinase-substrate interactions and aid in the design of inhibitor-based follow-up perturbation experiments. An interactive web interface allows investigation of predicted kinase-substrate interactions from human and major eukaryotic model organisms.

Let's start with example sequences

The web service allows a user to submit protein sequence, name, or id.

Click

Let's start with example sequences

http://kinomexplorer.info/
http://networkin-beta.cbs.dtu.dk/
NetworKIN

Set your organism
Human

Paste sequences (FASTA format) or protein names (one per line) below
Example: #1, #2, #3
In case you have more than 100 sequences, please consider using the high throughput workflow.

Select Phosphosites

Click
NetworKIN suggests proteins based on sequence homology

The submitted sequences mapped to multiple targets in ENSEMBL (e-value <1e-40, identity >90%) Please specify the right one.

Click after confirming that the suggested proteins are right ones
You can choose phosphorylation sites, otherwise the server takes all S, T, Y sites as candidates.
<table>
<thead>
<tr>
<th>Protein ID</th>
<th>Description</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q9Z2H5</td>
<td>ENSP0000037168</td>
<td>806</td>
</tr>
<tr>
<td>O00418</td>
<td>ENSP00000263026</td>
<td>80</td>
</tr>
<tr>
<td>O00151</td>
<td>ENSP00000360305</td>
<td>720</td>
</tr>
<tr>
<td>Q9Z2Q6</td>
<td>ENSP00000328190</td>
<td>80</td>
</tr>
<tr>
<td>Q9Z266</td>
<td>ENSP00000371763</td>
<td>720</td>
</tr>
</tbody>
</table>

The page at kinomexplorer.info says:
No sites for prediction selected
Should the prediction run on all sites?

Choose: 
- Cancel
- OK

Click on 'OK'.
**Results**

Residues that are predicted to be phosphorylated:

<table>
<thead>
<tr>
<th>Residue</th>
<th>Kinase</th>
<th>Score</th>
<th>Phosphorylation Site</th>
</tr>
</thead>
<tbody>
<tr>
<td>S18</td>
<td>MAPK13</td>
<td>2.0509</td>
<td>VDGQQSPRAGH</td>
</tr>
<tr>
<td>S74</td>
<td>MAPK13</td>
<td>2.0509</td>
<td>YSSGSQPANSF</td>
</tr>
<tr>
<td>T254</td>
<td>MAPK13</td>
<td>2.0168</td>
<td>DNIRLTPQAFS</td>
</tr>
<tr>
<td>S359</td>
<td>MAPK13</td>
<td>2.0168</td>
<td>EEKCGSPQVRT</td>
</tr>
<tr>
<td>S366</td>
<td>p70S6K</td>
<td>2.8915</td>
<td>QVTLSGSRPP</td>
</tr>
<tr>
<td>S396</td>
<td>MAPK13</td>
<td>2.0509</td>
<td>DLSQPSSAT</td>
</tr>
<tr>
<td>T401</td>
<td>MAPK13</td>
<td>2.0509</td>
<td>SPSSATPHSQK</td>
</tr>
<tr>
<td>S464</td>
<td>p70S6K</td>
<td>2.5262</td>
<td>HGHSYSNKRKYE</td>
</tr>
<tr>
<td></td>
<td>p70S6K</td>
<td>2.3401</td>
<td>HGHSYSNKRKYE</td>
</tr>
</tbody>
</table>

Kinases that are predicted to phosphorylate the site:

- MAPK13
- PKCbeta

**NetworKIN**

Minimum score: 2.00
Max. difference: 1.00
Min. # of Predictions: 2
Real time filter for results: 1

[Additional notes and data analysis from NetworKIN interface]
Exercise 1.

Q. Using NetworKIN, find top 3 kinases that are predicted to phosphorylate S39 position in human N-RAS protein.

Protein sequence of human N-RAS

>sp|P01111|RASN_HUMAN GTPase NRas OS=Homo sapiens GN=NRAS PE=1 SV=1
MTEYKLVVVGAGVSGKSAITQIQLQNHQVDEYDPTIEDSYRQVVIDGETCCLLDIILDTAG
QEEYSAMRDQYMTGEGFLCVAINEQSKSFADINLYREQQIKRVSDDVPMVVGKNKCDL
PRTVDTKQAHELAKSYGIPFIETSATROQGLVEDAFYTLVIREQRMMKNLSSDDGTQG
CMGLPCVVM

NetworKIN web site

http://kinomexplorer.info/
http://networkin-beta.cbs.dtu.dk/
Not all the residues are accessible by kinases.

Core residues are not accessible by kinases.

http://pipe.scs.fsu.edu/
Exercise 2. Predict phosphosites of MDM2 (human) and kinases that phosphorylate them.

- Find the sequence at Uniprot.org.
- Use NetPhos score cutoff 0.7 for phosphosite prediction
- Use NetworKIN score cutoff 20 for kinase prediction

How many phosphosite-kinase interactions were predicted?
Which one is the most probable one?

Exercise 3. Add S395 manually and run again with the same cutoffs.

How many phosphosite-kinase interactions were predicted?
Which one is the most probable one?
High-throughput workflow of NetworKIN

Sample

Thousands of phosphosites on thousands of proteins

You are only interested in the detected phosphosites, not all S, T, Y sites
Set your organism

[Human]

Paste sequences (FASTA format) or protein names (one per line) below

In case you have more than 100 sequences, please consider using the [high throughput workflow](#).

[Select Phosphosites]

Click
**Select the sequence database:**

Ensembl 74 (human)

**Put modified sites below in the format Protein ID, Position, Residue (space/tab delimited) Example:#1, #2, ProteomeDiscoverer**

<table>
<thead>
<tr>
<th>Protein ID</th>
<th>Position</th>
<th>Residue</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENSP00000298139</td>
<td>1292S</td>
<td></td>
</tr>
<tr>
<td>ENSP00000325863</td>
<td>678 S</td>
<td></td>
</tr>
<tr>
<td>ENSP00000265433</td>
<td>278 S</td>
<td></td>
</tr>
<tr>
<td>ENSP00000265433</td>
<td>343 S</td>
<td></td>
</tr>
<tr>
<td>ENSP00000298139</td>
<td>1141S</td>
<td></td>
</tr>
<tr>
<td>ENSP00000269305</td>
<td>378 S</td>
<td></td>
</tr>
<tr>
<td>ENSP00000325863</td>
<td>264 S</td>
<td></td>
</tr>
<tr>
<td>ENSP00000265433</td>
<td>58 S</td>
<td></td>
</tr>
<tr>
<td>ENSP00000278616</td>
<td>440 S</td>
<td></td>
</tr>
</tbody>
</table>

**OR select a MaxQuant output file:**

Choose File: No file chosen

---

**Raw output from the search engine**
Choose the right sequence database and version

Select the sequence database:

- Ensembl 74 (human)
- Ensembl 73 (human)
- Ensembl 72 (human)
- Ensembl 71 (human)
- Ensembl 70 (human)
- Ensembl 69 (human)
- Ensembl 68 (human)
- Uniprot 2013/12 (human)
- NCBI 62 (human)
- Ensembl 74 (yeast)
- Ensembl 73 (yeast)
- Ensembl 72 (yeast)
- Ensembl 71 (yeast)
- Ensembl 70 (yeast)
- Ensembl 69 (yeast)
- Ensembl 68 (yeast)
- Uniprot 2013/12 (yeast)
- NCBI 62 (yeast)
- Uniprot (Maxquant 01/2013)
- STRING 9.0

This is the high-throughput interface for NetworKIN. Click here to go back to the low-throughput version.
Not all submitted identifier could not be mapped to a unique STRING entity. Please select a desired isotype or homolog.
Not all submitted identifier could not be mapped to a unique STRING entity. Please select a desired isotype or homolog.

- **ENSP00000355746**
  - PSEN2 (100.00% 448AA)
  - PSEN1 (64.81% 425AA)
- **ENSP00000352925**
  - TRRAP (99.22% 3859AA)
  - ATR (24.02% 226AA)
- **ENSP00000402030**
  - LIG4
- **ENSP00000366863**
  - TBC1D4
- **ENSP00000269305**
  - TP53
- **ENSP00000298139**
  - WRN
- **ENSP00000347232**
  - BLM
- **ENSP00000278616**
  - ATM
- **ENSP00000295266**
  - PDHA2
- **ENSP00000358576**
  - DCLRE1B
<table>
<thead>
<tr>
<th>Gene</th>
<th>Kinase</th>
<th>Score</th>
<th>Phospho Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENSP00000219476</td>
<td>S1341 PDHK1</td>
<td>8.24</td>
<td>SSVSVSVQEKS</td>
</tr>
<tr>
<td></td>
<td>GSK3beta</td>
<td>5.77</td>
<td>SSVSVSVQEKS</td>
</tr>
<tr>
<td>ENSP00000261716</td>
<td>S554 ATM</td>
<td>128.36</td>
<td>NSFLESLKREY</td>
</tr>
<tr>
<td></td>
<td>S990 ATM</td>
<td>95.10</td>
<td>STSVTNISNG</td>
</tr>
<tr>
<td>ENSP00000263710</td>
<td>S1196 CK1delta</td>
<td>8.02</td>
<td>KFSFRSQEDLN</td>
</tr>
<tr>
<td></td>
<td>CK1epsilon</td>
<td>5.16</td>
<td>KFSFRSQEDLN</td>
</tr>
<tr>
<td>ENSP00000265335</td>
<td>S635 ATM</td>
<td>167.80</td>
<td>FDVCGSDFSES</td>
</tr>
<tr>
<td></td>
<td>BRCT BRCA1</td>
<td>3.98</td>
<td>FDVCGSDFSES</td>
</tr>
<tr>
<td>ENSP00000265433</td>
<td>S58 ATM</td>
<td>228.14</td>
<td>SVTNLSTDEI</td>
</tr>
<tr>
<td></td>
<td>S278 ATM</td>
<td>228.14</td>
<td>TGITNSQLIP</td>
</tr>
<tr>
<td></td>
<td>BRCT BRCA1 MDC1</td>
<td>3.74</td>
<td>TGITNSQLIP</td>
</tr>
<tr>
<td></td>
<td>MDC1</td>
<td>3.42</td>
<td>TGITNSQLIP</td>
</tr>
<tr>
<td></td>
<td>S343 ATM</td>
<td>228.14</td>
<td>PGPSLSGVSV</td>
</tr>
<tr>
<td></td>
<td>BRCT BRCA1 MDC1</td>
<td>3.74</td>
<td>PGPSLSGVSV</td>
</tr>
<tr>
<td></td>
<td>MDC1</td>
<td>3.42</td>
<td>PGPSLSGVSV</td>
</tr>
<tr>
<td></td>
<td>S397 ATM</td>
<td>214.52</td>
<td>KFRMLSDAPT</td>
</tr>
<tr>
<td></td>
<td>S615 ATM</td>
<td>213.95</td>
<td>ESSKISSENEI</td>
</tr>
<tr>
<td>ID</td>
<td>Gene</td>
<td>Score</td>
<td>Peptide</td>
</tr>
<tr>
<td>-------------</td>
<td>------------</td>
<td>-------</td>
<td>-------------------------------</td>
</tr>
<tr>
<td>ENSP00000219476</td>
<td>Kinase ATM</td>
<td>128.36</td>
<td>NSFLESKREY</td>
</tr>
<tr>
<td>ENSP00000261716</td>
<td>Kinase ATM</td>
<td>95.10</td>
<td>STSVTSISNG</td>
</tr>
<tr>
<td>ENSP00000263710</td>
<td>Kinase ATM</td>
<td>228.14</td>
<td>SVTNLSITDEI</td>
</tr>
<tr>
<td>ENSP00000265335</td>
<td>Kinase ATM</td>
<td>214.52</td>
<td>KFRMLSCADAPT</td>
</tr>
<tr>
<td>ENSP00000265433</td>
<td>Kinase ATM</td>
<td>213.95</td>
<td>ESSKISGNEEI</td>
</tr>
<tr>
<td>ENSP00000269305</td>
<td>Kinase ATM</td>
<td>167.80</td>
<td>SSVPSKMTQ</td>
</tr>
<tr>
<td>ENSP00000278616</td>
<td>Kinase ATM</td>
<td>228.14</td>
<td>KQGSTSHKKL</td>
</tr>
<tr>
<td>ENSP00000288986</td>
<td>Kinase ATM</td>
<td>151.88</td>
<td>QSYTTTRESS</td>
</tr>
<tr>
<td>ENSP00000295266</td>
<td>Kinase ATM</td>
<td>130.14</td>
<td>KAYSSGCPIS</td>
</tr>
<tr>
<td>ENSP00000298139</td>
<td>Kinase ATM</td>
<td>130.14</td>
<td>IGMHSLAVKA</td>
</tr>
</tbody>
</table>
Click
Exercise 4. Predict ATM kinase targets with example #1 through high-throughput workflow. (Use a cutoff: 20)

How many phosphosite-ATM kinase interactions were predicted? Which one is the most probable one?
Exercise 5. Predict phospho-interactions using known phosphosites (phosphosites.txt) through high-throughput workflow. (Choose Ensembl 70 (human) and Use a cutoff: 20)

• How many phosphosite-ATM kinase interactions were predicted?
• Which one is the most probable one?
• Are the substrate and kinase involved in the same biological process?
• What happens if you choose Ensembl 68 (human) instead of Ensembl 70
Exercise 6. Predict kinases that phosphorylate T1162 and T2446 of human MTOR protein through both Low- and High-throughput workflows.